DOCUMENT RESUME

ED 240 144 TM 840 058

AUTHOR Roth, Rod
TITLE Relationships among NTE Cut-Scores, Not Valid Items, NTE Tests and Curriculum Match and Minimally Competent Examinees.
PUB DATE 16 Nov 83
NOTE 8p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (12th, Nashville, TN, November 16, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Faculty; *Cutting Scores; Evaluation Criteria; Higher Education; *Item Analysis; Teacher Education Curriculum; *Teacher Evaluation; Test Items; *Test Validity
IDENTIFIERS *Arkansas; *National Teacher Examinations

ABSTRACT

College faculty (n=171) from 16 Arkansas colleges were asked to make validity and cut score judgments about the test items for the 1982 Arkansas National Teacher Examination (NTE) study of 23 area examinations. Each of the 23 data collection panels began with a training session which included specific directions for the estimates of the judges. Results indicate that (1) the closer the test-curriculum match, the greater is the likelihood that the test has more valid items, (2) the more items not valid on a test, the higher would be the percent of those who would score lower than the minimally competent examinee, (3) the greater the match between items and curriculum content, the higher the derived cut-scores, (4) the lower the expected failure rate, the higher the derived cut-score, and (5) the greater the match between test items and curriculum content, the lower the expected failure rate. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED240144

RELATIONSHIPS AMONG NTE CUT-SCORES, NOT VALID ITEMS, NTE TESTS AND CURRICULUM MATCH AND MINIMALLY COMPETENT EXAMINEES

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Dr. Rod Roth, Dean
College of Education
The University of Alabama

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. Roth

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 840 058

A paper presented at the Annual Meeting of the Mid-South Educational Research Association in Nashville, Tennessee, on November 16, 1983

The purpose of this paper is to present some relationships among several variables that were collected from college faculty during the Arkansas National Teacher Examination (NTE) validation and cut-score study. The data were collected in April, 1982. The specific research questions for this paper are:

1. What is the relationship between the number of not valid items (variable 1) and the median percent of items on the NTE area examination covered by the preparation curriculum (variable 3)?

2. What is the relationship between the number of not valid items (variable 1) and the median percent who might be expected to score lower than the minimally competent examinee (variable 4)?

3. What is the relationship between the median percent of items on the area examination covered by the curriculum (variable 3) and the derived cut-score (variable 2)?

4. What is the relationship between the median percent who might be expected to score lower than the minimally competence examinee (variable 4) and the derived cut-score (variable 2)?

5. What is the relationship between (variable 3) and (variable 4)?

Methodology

This section of the paper presents the judge selection procedures, data collection instruments, and data analysis procedures.

Judge Selection

Each College of Education dean in the sixteen teacher training institutions was asked to nominate judges from his/her institution for the NTE study. Each dean was asked to nominate judges only in the NTE areas in which the college had approved

certification programs. The nominated judges filled out a nomination form which included information about race, sex, years of teaching experience and courses taught.

The actual selection of the final set of judges was made by this writer from the pool of nominations made by the deans. A panel of judges was selected for each NTE area examination.

Some of the criteria used to select the judges were race, sex, years of teaching/administrative experience, teaching assignment and for college faculty, the numbers of graduates produced by their institution.

A total of 171 faculty from 16 Arkansas colleges were used as judges for the NTE study of 23 area examinations. The average number of faculty on the 23 different judging panels was seven. A total of 161 practitioners from Arkansas public schools were also used as judges in the study. They were not, however, asked to respond to several of the variables used in this paper.

Data Collection

Each data collection session began with a training session. It included a legal history of the NTE in Arkansas, purpose of the NTE area examinations, the need for state validation, and the NTE study design including how the judges were selected. The training session also included very specific directions for the validity and cut-score judgments. The directions were:

> The first rating you will make concerns Item Relevance. This will be used for test validation. In order to make this judgment, you should read the item, the "correct" answer, and the distractors. (The correct answer is underlined in the test booklet.) You should then judge the relevance of the content measured by the question with respect to the domain of knowledge you believe a minimally qualified entry-level person in the certification area should possess.

If you believe the content of the question is irrelevant to the domain of knowledge a minimally qualified entry-level person in this field should possess, then you should fill-in circle 1 on your answer sheet in the Relevance column to signify "Not Relevant."

If you believe the content of the question is of doubtful or questionable relevance to the domain of knowledge a minimally qualified entry-level person in this field should possess, then you should fill-in circle 2 on your answer sheet in the Relevance column to signify "Questionable."

If you believe the content of the question is important, but not quite crucial, to the domain of knowlege a minimally qualified entry-level person in this field should possess, then you should fill-in circle 3 on your answer sheet in the Relevance column to signify "Important."

If you believe the content of the question is of crucial importance to the domain of knowledge a minimally qualified entry-level person in this field should possess, then you should fill-in circle 4 on your answer sheet in the Relevance column to signify "Crucial."

The second judgment you will make about each item will help determine the cut-score. You should imagine a hypothetical person, who in your judgment, has the minimum amount of academic knowledge to complete the preparation program required for certification in Arkansas and has the minimum amount of knowledge to perform in the field designated by the NTE area test. With this hypothetical person in mind, you are to estimate the probability that this minimally competent person would know the answer to the NTE item without guessing. Another way of thinking about this estimation process is to think of a group of minimally competent persons and then estimate the percent of minimally competent persons who would answer the NTE item correctly without guessing.

Before you make your estimate about the item, you should also realize the item difficulty based on the NTE norm group for the item. The item difficulty or the percent who have passed the item is written beside the item in the booklet.

You should mark your estimate for each item on the response sheet under the Probability column. You should use the following scale for these estimates:

Fill in circle 1, if your estimate is between .00 - .10
Fill in circle 2, if your estimate is between .11 - .20
Fill in circle 3, if your estimate is between .21 - .30
Fill in circle 4, if your estimate is between .31 - .40
Fill in circle 5, if your estimate is between .41 - .50
Fill in circle 6, if your estimate is between .51 - .60
Fill in circle 7, if your estimate is between .61 - .70
Fill in circle 8, if your estimate is between .71 - .80
Fill in circle 9, if your estimate is between .81 - .90
Fill in circle 10, if your estimate is between .91 - 1.0

After the faculty judges had made their judgments about each NTE item, they were asked additional questions. One of the questions (variable 3) was: Please indicate the approximate percent of items in this test that measure content covered in the preparation curriculum at your institution for this certification area. The other question (variable 4) was: Approximately what percent of the examinees from Arkansas preparation programs might be expected to score lower than the minimally competent examinee you had in mind as you evaluated the test items?

Data Analysis

The validity of each item was determined by computing an item mean for each item on the relevance scale. This scale had a range from one (Not Relevant) to four (Crucial). In order for an item to be considered valid, the mean score on the relevance scale had to be greater than 2.5. In other words, the item had to be rated by the judges as closer to the important category than to the questionable category. If half of the judges had rated the item questionable and the other half had rated the item important, then the item would not have met the validity criterion since the mean rating would have been 2.50. The number of not valid items (variable 1) was simply the total number of items for an area examination that did not meet the validity criterion.

The cut-score for each area examination was determined by a slight modification of a procedure known as the Angoff method. The first step for determining the cut-score was to determine an item mean on the probability scale. Since the judge had responded to a probability range for each item, the mid-point of the range was used to compute the item mean. For example, a one on the probability scale was converted to .05 since one represented the probability between .0 thru .1.

The raw score cut-score for each area examination was computed by summing the mean probabilities for only the items that had met the validity criterion. A conversion formula was used to convert the raw scores to NTE standard scores or derived scores (variable 2).

Variables three and four were determined by computing a median score from the range of scores on each of the two questions which were asked the college faculty judges.

Pearson correlation coefficients, with a N of 23, were computed to determine the five relationships posed by the five research questions. In other words, judgments from each of the 23 NTE area examinations yielded four variables per examination.

Results

This section of the paper presents the results and a brief interpretation of the results. The results were:

Question 1, Variables 1-3 $r = -.61$ $p = .001$

Question 2, Variables 1-4 $r = .37$ $p = .04$

Question 3, Variables 2-3 $r = .62$ $p = .001$

Question 4, Variables 2-4 $r = -.53$ $p = .005$

Question 5, Variables 3-4 $r = -.54$ $p = .004$

The significant negative correlation for question one indicates that the greater the test content is covered in the preparation curriculum the fewer the number of items considered not valid on the NTE area examinations. In other words, the closer the test-curriculum match, the likelihood is greater that the test had more valid items.

The significant positive correlation for question two indicates that the more not valid items on a test, the higher the percent would be who would score lower than the minimally competent examinee. Another interpretation is that the more valid the test, the less likely for student failure.

The significant positive correlation for question three indicates that the greater the match between items and curriculum content, the higher the derived cut-scores.

The significant negative correlation for question four indicates that the lower the expected failure rate, the higher the derived cut-score. Stated another way, the lower the derived cut-scores, the higher the expected failure rate.

The significant negative correlation for question five indicates that the greater the match between test items and curriculum content, the lower the expected failure rate. In other words, when faculty felt the tests matched the curriculum, they also felt that the failure rates would be low.

In conclusion, it is difficult to provide precise conclusions from this study because the writer did not present hypotheses. The writer instead chose to ask some interesting questions concerning four different judgments made by college faculty. I do, however, feel that the relationships can lead to theory building in the fields of standard setting and validation studies.